Efficient, Nearly Optimal Nonlinear Scaling of Predictors for Statistical Forecast Equations

F. Wesley Wilson
MIT Lincoln Laboratory
244 Wood St.
Lexington, MA 02173
phone (781) 893-9324 fax (781) 981-0632 email wesw@ll.mit.edu

Award Number: N0001499MP0046
http://www.onr.navy.mil/sci_tech/ocean/onrpgabr.htm

LONG-TERM GOALS

My long term goal is to develop a methodology for the determination of statistical forecast models, which includes data-determined nonlinear data scaling, which can be applied to data sets of moderate size, and for which the development process can be substantially automated.

OBJECTIVES

The objectives for this year are the development of manual techniques for the selection of piecewise linear structure functions, to evaluate their utility for the development of statistical forecast models, and to evaluate their value for the improvement of statistical forecast equations.

These models will be applied to several cases for the development of forcast equations for the clearing of marine stratus at San Francisco International Airport. Of particular interest is the forecast equations based on satelite data.

APPROACH

I will investigate a modification of the General Additive Model (GAM) approach to the development of nonlinear scaling of predictors. This approach seeks piecewise linear structure functions with a small number of knots. These models are determined by nonlinear optimization, where the objective function is the correlation coefficent of the predictor and the predictand. Since the number of degrees of freedom in this approach is substantially smaller than is required for traditional GAM or neural net developments, the process can be applied with considerably smaller training sets. Since the objective function is determined independently for each predictor, there is also the potential to apply SVD techniques for optimal predictor selection. This provides an additional advantage over GAM, where a change in the predictor list can modify the scaling of all of the predictors. Finally, these simpler stuctures lend themselves to an automated scaling strategy. A prototype automated scaling code has been developed, but has not yet been fully evaluated.

For the development of statistical forecast equations, three approaches are considered:

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **30 SEP 2001**	2. REPORT TYPE	3. DATES COVERED **00-00-2001 to 00-00-2001**

4. TITLE AND SUBTITLE **Efficient, Nearly Optimal Nonlinear Scaling of Predictors for Statistical Forecast Equations**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **MIT Lincoln Laboratory,244 Wood St.,,Lexington,,MA,02173**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The objectives for this year are the development of manual techniques for the selection of piecewise linear structure functions, to evaluate their utility for the development of statistical forecast models and to evaluate their value for the improvement of statistical forecast equations.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**4**	

Multi-Linear Regression provides forecast equations of the form

$$F = W_0 + \Sigma\, W_i\, P_i$$

The Generalized Additive Model approach provides forecast equations of the form

$$F = W_0 + \Sigma\, W_i\, f_i(P_i)$$

where the structure functions f_i are determined by the training data during the regression.

The modified Generalized Additive Model approach provides forecast equations of the form

$$F = W_0 + \Sigma\, W_i\, f_i(P_i)$$

where the structure functions f_i are determined by the training data, but prior to the regression.

WORK COMPLETED

We have completed two studies this year:

1. The development of a spread sheet technique for the selection of optimal piecewise linear structure functions.

2. The application of this technique to the development of statistical forecast equations for the clearing of marine stratus at San Francisco International Airport.

RESULTS

This technique has been applied for the development of two classes of forecast equations for the clearing of marine stratus at San Francisco International Airport. In the first case, the predictors are taken from measurements of the boundary layer structure near the airport, such as the height of the inversion base, the height of the ceiling, and the wind components. In the second case, the predictors are taken from statistical measures of clusters of visible satellite data, where the clusters are associated with a tiling of the region into meteorologically significant sectors.

In each case, forecast equations for the time of clearing are developed against conditional climatology, where the condition is that stratus is present at the time the forecast is issued. In every case, the modified GAM approach provided an improvement of the forecast equations over simple multilinear regression. One estimate of the magnitude of the improvement is the reduction of the mean absolute error (MAE). By this measure, the error reduction ranged from 6% to 15% compared with multilinear regression. Since the training data set is too limited for the application of GAM, it is not possible to make a direct comparison with the models that might be developed using traditional GAM techniques. However, these results are similar to the improvements, which are typically achieved in GAM applications.

IMPACT/APPLICATIONS

An important application of this technology would be to develop statistical forecast equations for 2-4 hour forecasts of ceiling and cloud evolution in regions that do not have a substantial local surface historical data archive. In this application, forecast equations could be developed based on archived satellite data, providing an opportunity to rapidly develop statistical forecast aids for regions that are unexpectedly thrust to the forefront.

TRANSITIONS

These techniques are being applied for the development of forecast products for use by the National Weather Service at San Francisco International Airport.

RELATED PROJECTS

The development and evaluation of these forecast aids is partially supported by the FAA' s Aviation Weather Research Program for the forecast of the dissipation of low clouds at San Francisco International Airport.

REFERENCES

Vislocky, R.V., and J.M. Fritsch, 1995: Generalized Additive Models versus Linear Regression in Generating Probabilistic MOS Forecasts of Aviation Weather Parameters. *Weather and Forecasting*, **10**, pp. 669-680.

Wilson, F. W. and D. A. Clark, 2000: Forecast aids to lessen the impact of marine stratus on San Francisco International Airport**,** Preprints, Ninth Conference on Aviation, Range, and aerospace Meteorology, AMS, Orlando.